

A simple alternative to the h-index

Ludo Waltman and Nees Jan van Eck

| ERIM REPORT SERIES <i>RESEARCH IN MANAGEMENT</i> | |
|--|--|
| ERIM Report Series reference number | ERS-2009-043-LIS |
| Publication | July 2009 |
| Number of pages | 5 |
| Persistent paper URL | http://hdl.handle.net/1765/16556 |
| Email address corresponding author | lwaltman@ese.eur.nl |
| Address | Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl |

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES *RESEARCH IN MANAGEMENT*

| ABSTRACT AND KEYWORDS | |
|-----------------------|--|
| Abstract | The h-index is a popular bibliometric performance indicator. We discuss a fundamental problem of the h-index. We refer to this problem as the problem of inconsistency. There turns out to be a very simple bibliometric indicator that has similar properties as the h-index and that does not suffer from the inconsistency problem. We argue that the use of this indicator is preferable over the use of the h-index. |
| Free Keywords | bibliometric indicator, h-index, highly cited publications, consistency |
| Availability | The ERIM Report Series is distributed through the following platforms: Academic Repository at Erasmus University (DEAR), DEAR ERIM Series Portal Social Science Research Network (SSRN), SSRN ERIM Series Webpage Research Papers in Economics (REPEC), REPEC ERIM Series Webpage |
| Classifications | The electronic versions of the papers in the ERIM report Series contain bibliographic metadata by the following classification systems: Library of Congress Classification, (LCC) LCC Webpage Journal of Economic Literature, (JEL), JEL Webpage ACM Computing Classification System CCS Webpage Inspec Classification scheme (ICS), ICS Webpage |

A simple alternative to the h -index

Ludo Waltman

Nees Jan van Eck

Econometric Institute, Erasmus School of Economics

Erasmus University Rotterdam

P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

E-mail: {lwaltman,nvaneck}@ese.eur.nl

Centre for Science and Technology Studies, Leiden University

P.O. Box 905, 2300 AX Leiden, The Netherlands

Abstract

The h -index is a popular bibliometric performance indicator. We discuss a fundamental problem of the h -index. We refer to this problem as the problem of inconsistency. There turns out to be a very simple bibliometric indicator that has similar properties as the h -index and that does not suffer from the inconsistency problem. We argue that the use of this indicator is preferable over the use of the h -index.

Keywords

Bibliometric indicator, h -index, highly cited publications, consistency.

1 Introduction

Within a short period of time, the h -index proposed by Hirsch (2005) has become a very popular bibliometric performance indicator. The following properties of the h -index seem to be the main reason for its popularity:

- The h -index is relatively insensitive to large numbers of lowly cited publications.
- The h -index is relatively insensitive to a few very highly cited publications.

- The calculation of the h -index is easy to understand.

Researchers have also identified various weak points of the h -index. Most of these weak points can be dealt with relatively easily without the need to abandon the basic ideas underlying the h -index. This has resulted in a large number of h -index variants. In this note, we focus on a more fundamental problem of the h -index. We refer to this problem as the problem of inconsistency. The problem of inconsistency cannot be solved by introducing yet another variant of the h -index. In fact, the problem is shared by all h -index variants. However, there turns out to be a very simple (and well-known) indicator that has similar properties as the h -index and that does not suffer from the inconsistency problem. We argue that the use of this indicator is preferable over the use of the h -index.

2 The problem of inconsistency

To illustrate the problem of inconsistency, consider the following example. Suppose that researchers A and B both have four publications. Researcher A has three publications with five citations each and one publication with zero citations. Researcher B has four publications with four citations each. Suppose next that researchers A and B both achieve the same performance improvement. More specifically, suppose that they both obtain two new publications with five citations each. Researcher A's h -index then increases from three to five, while researcher B's h -index remains equal to four. This means that the ranking of researchers A and B relative to each other has reversed. Initially researcher A was ranked below researcher B, but after adding the new publications the situation is exactly the other way around. We regard this as a highly undesirable outcome. Researchers A and B have both achieved the same performance improvement, but despite of that their ranking relative to each other has reversed. In our opinion, this is unnatural and very difficult to justify.

Because of the above problem, we call the h -index an inconsistent indicator (for a formal definition of the notion of consistency, see Waltman & Van Eck, 2009). The inconsistency problem of the h -index has also been recognized by Marchant (2009a, 2009b) (see also Rousseau, 2008). Marchant (2009a, p. 335) concludes that because of the inconsistency problem “the ranking based on the h -index is in many circumstances probably not reasonable”. We fully agree with this conclusion.

The following example provides another illustration of the inconsistency problem of the h -index. Suppose that the h -index is used to compare the performance of two research groups, research group A and research group B. Research group A consists of five researchers. Each researcher in research group A has five publications with five citations each. Hence, in total research group A has 25 publications with five citations each. Research group B also consists of five researchers. Each researcher in research group B has two publications with ten citations each. Hence, in total research group B has ten publications with ten citations each. It is clear that each researcher in research group A has an h -index of five while each researcher in research group B has an h -index of two. This means that according to the h -index each researcher in research group A outperforms each researcher in research group B. Based on this result, it seems natural to expect that research group A as a whole outperforms research group B as a whole. However, this is not the case. Research group A has an h -index of five, while research group B has an h -index of ten. Hence, the h -index indicates that research group A is outperformed by research group B rather than the other way around. This means that the h -index calculated at the level of research groups contradicts the h -index calculated at the level of individual researchers. We regard this as a rather odd result.

3 An alternative indicator

To avoid the problem of inconsistency, we propose a simple alternative to the h -index. This alternative is the number of highly cited publications, that is, the number of publications for which the number of citations exceeds a certain threshold. The number of highly cited publications is a well-known indicator (e.g., Plomp, 1990) that has a number of similar properties as the h -index, namely insensitivity to large numbers of lowly cited publications, insensitivity to a few very highly cited publications, and an easy-to-understand calculation. There is one very important difference between the highly cited publications (HCP) indicator and the h -index. This difference is that unlike the h -index the HCP indicator does not suffer from the problem of inconsistency (Marchant, 2009a; Waltman & Van Eck, 2009). Because of this difference, we believe that the use of the HCP indicator is preferable over the use of the h -index.

In his proposal for the h -index, Hirsch (2005) also discusses the possibility of using the number of highly cited publications as a bibliometric performance indicator. According to

Hirsch, the HCP indicator has a significant disadvantage. This disadvantage is that the threshold for determining what counts as highly cited and what does not “is arbitrary and will randomly favor or disfavor individuals” (Hirsch, 2005, p. 16569). We agree with Hirsch that this is a weak point of the HCP indicator. However, unlike what Hirsch claims, the h -index has a similar weak point. As we all know, the h -index is defined as follows: A researcher has h -index h if h of his n publications have at least h citations each and the other $n - h$ publications have fewer than $h + 1$ citations each. This definition involves some arbitrariness because the h -index could equally well have been defined in, for example, the following way: A researcher has h -index h if h of his n publications have at least $2h$ citations each and the other $n - h$ publications have fewer than $2(h + 1)$ citations each. Or the following definition of the h -index could have been given: A researcher has h -index h if h of his n publications have at least $h/2$ citations each and the other $n - h$ publications have fewer than $(h + 1)/2$ citations each. Hirsch does not provide any argument why his definition of the h -index would be better than alternative definitions such as those given above. Because of this, the way in which Hirsch defines the h -index is somewhat arbitrary (Van Eck & Waltman, 2008). The arbitrariness of the definition of the h -index is also pointed out by Lehmann, Jackson, and Lautrup (2006, 2008), who note that the h -index is based on a comparison of two quantities that have different units (publications vs. citations). Because not only the use of the HCP indicator involves arbitrariness but the use of the h -index does so as well, we do not consider arbitrariness an important factor in the choice between the h -index and the HCP indicator.

4 Conclusion

The h -index and the HCP indicator have quite similar properties. However, the h -index suffers from a fundamental problem, namely the problem of inconsistency. Because the HCP indicator does not suffer from this problem, we consider the use of this indicator preferable over the use of the h -index. The use of the HCP indicator indeed involves some arbitrariness, as noted by Hirsch (2005), but the same holds for the use of the h -index. Arbitrariness is therefore not a valid argument for rejecting the HCP indicator in favor of the h -index.

References

- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.
- Lehmann, S., Jackson, A. D., & Lautrup, B. E. (2006). Measures for measures. *Nature*, 444, 1003–1004.
- Lehmann, S., Jackson, A. D., & Lautrup, B. E. (2008). A quantitative analysis of indicators of scientific performance. *Scientometrics*, 76(2), 369–390.
- Marchant, T. (2009a). An axiomatic characterization of the ranking based on the h -index and some other bibliometric rankings of authors. *Scientometrics*, 80(2), 327–344.
- Marchant, T. (2009b). Score-based bibliometric rankings of authors. *Journal of the American Society for Information Science and Technology*, 60(6), 1132–1137.
- Plomp, R. (1990). The significance of the number of highly cited papers as an indicator of scientific prolificacy. *Scientometrics*, 19(3–4), 185–197.
- Rousseau, R. (2008). Luckily, science focuses on achievements. Some thoughts related to the h -index. *ISSI Newsletter*, 4(3), 49–50.
- Van Eck, N. J., & Waltman, L. (2008). Generalizing the h - and g -indices. *Journal of Informetrics*, 2(4), 263–271.
- Waltman, L., & Van Eck, N. J. (2009). A taxonomy of bibliometric performance indicators based on the property of consistency. In *Proceedings of the 12th International Conference on Scientometrics and Informetrics* (pp. 1002–1003).

Publications in the Report Series Research* in Management

ERIM Research Program: “Business Processes, Logistics and Information Systems”

2009

How to Normalize Co-Occurrence Data? An Analysis of Some Well-Known Similarity Measures

Nees Jan van Eck and Ludo Waltman

ERS-2009-001-LIS

<http://hdl.handle.net/1765/14528>

Spare Parts Logistics and Installed Base Information

Muhammad N. Jalil, Rob A. Zuidwijk, Moritz Fleischmann, and Jo A.E.E. van Nunen

ERS-2009-002-LIS

<http://hdl.handle.net/1765/14529>

Open Location Management in Automated Warehousing Systems

Yugang YU and René B.M. de Koster

ERS-2009-004-LIS

<http://hdl.handle.net/1765/14615>

VOSviewer: A Computer Program for Bibliometric Mapping

Nees Jan van Eck and Ludo Waltman

ERS-2009-005-LIS

<http://hdl.handle.net/1765/14841>

Nash Game Model for Optimizing Market Strategies, Configuration of Platform Products in a Vendor Managed Inventory (VMI) Supply Chain for a Product Family

Yugang Yu and George Q. Huang

ERS-2009-009-LIS

<http://hdl.handle.net/1765/15029>

A Mathematical Analysis of the Long-run Behavior of Genetic Algorithms for Social Modeling

Ludo Waltman and Nees Jan van Eck

ERS-2009-011-LIS

<http://hdl.handle.net/1765/15181>

A Taxonomy of Bibliometric Performance Indicators Based on the Property of Consistency

Ludo Waltman and Nees Jan van Eck

ERS-2009-014-LIS

<http://hdl.handle.net/1765/15182>

A Stochastic Dynamic Programming Approach to Revenue Management in a Make-to-Stock Production System

Rainer Quante, Moritz Fleischmann, and Herbert Meyr

ERS-2009-015-LIS

<http://hdl.handle.net/1765/15183>

Some Comments on Egghe's Derivation of the Impact Factor Distribution

Ludo Waltman and Nees Jan van Eck

ERS-2009-016-LIS

<http://hdl.handle.net/1765/15184>

The Value of RFID Technology Enabled Information to Manage Perishables

Michael Ketzenberg, and Jacqueline Bloemhof

ERS-2009-020-LIS

<http://hdl.handle.net/1765/15412>

The Environmental Gains of Remanufacturing: Evidence from the Computer and Mobile Industry
J. Quariguasi Frota Neto, and J.M. Bloemhof
ERS-2009-024-LIS
<http://hdl.handle.net/1765/15912>

Economic Modeling Using Evolutionary Algorithms: The Effect of a Binary Encoding of Strategies
Ludo Waltman, Nees Jan van Eck, Rommert Dekker, and Uzay Kaymak
ERS-2009-028-LIS
<http://hdl.handle.net/1765/16014>

Language Selection Policies in International Standardization – Perception of the IEC Member Countries
Hans Teichmann and Henk J. de Vries
ERS-2009-031-LIS
<http://hdl.handle.net/1765/16038>

Dominant Design or Multiple Designs: The Flash Memory Card Case
Henk J. de Vries, Joost P.M. de Ruijter and Najim Argam
ERS-2009-032-LIS
<http://hdl.handle.net/1765/16039>

Standards Education Policy Development: Observations based on APEC Research
Donggeun Choi, Henk J. de Vries and Danbee Kim
ERS-2009-033-LIS
<http://hdl.handle.net/1765/16040>

Scheduling deliveries under uncertainty
Adriana F. Gabor, Rommert Dekker, Timon van Dijk, and Peter van Scheepstal
ERS-2009-040-LIS
<http://hdl.handle.net/1765/16236>

A simple alternative to the h-index
Ludo Waltman and Nees Jan van Eck
ERS-2009-043-LIS
<http://hdl.handle.net/1765/16556>

Disruption Management of Rolling Stock in Passenger Railway Transportation
Lars Kjaer Nielsen and Gabor Maroti
ERS-2009-046-LIS
<http://hdl.handle.net/1765/16557>

* A complete overview of the ERIM Report Series Research in Management:
<https://ep.eur.nl/handle/1765/1>

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship